

ROCm ai benchmark



Sign in

AI Mode

All

Videos

Images

News

Shopping

Short videos

More ▾

Tools ▾

## ✦ AI Overview



ROCm (Radeon Open Compute) is AMD's open-source software platform for GPU-accelerated computing, particularly for AI workloads. It provides tools and libraries to optimize and benchmark AI performance on AMD hardware. Key components include HIP (Heterogeneous-Compute Interface for Portability) for code portability, and various benchmarks and tools for assessing performance of AI models and systems. [↗](#)

Here's a breakdown of how ROCm is used for AI benchmarking:

### 1. System Health Checks and Validation:

#### **ROCm Validation Suite (RVS):**

RVS includes tests for GPU compute capabilities (FLOPS), memory bandwidth, and interconnect performance, crucial for validating hardware setup before running AI workloads.

#### **GPU Stress Tests:**

##### HIP (ROCm) semantics — PyTorch 2.7 documentation

ROCm™ is AMD's open source software platform for GPU-accelerated high performance computing and machine learning. HIP is ROCm's C+

 [PyTorch documentation](#) 

##### AMD ROCm™ Software for AI

Optimized for AI Workloads AMD ROCm™ is an open software stack offering a suite of optimizations for AI workloads and supporting t...

 [AMD](#) 

##### System health benchmarks - ROCm Documentation - AMD

May 13, 2025 — Before running AI workloads, it is important to validate that your AMD hardware is configured correctly and is perform...

 [ROCm Documentation](#) 


##### AMD's AI Chip Strategy: Gaining Ground in a Competitive Market

Sep 16, 2024 — AMD has developed a comprehensive suite of software




AMD provides tools and examples for benchmarking pre-training and fine-tuning of large language models (LLMs) using PyTorch on ROCm. 


### Inference Benchmarks:

ROCm offers tools and techniques for optimizing and benchmarking inference performance, including quantization, GEMM tuning, and composable kernels. 





### vLLM Benchmarking:

AMD provides guidance and tools for benchmarking online inference with vLLM on AMD GPUs, including instructions for setting up the environment and running the benchmarks. 

### MLPerf:

MLPerf is a benchmark suite developed by MLCommons to measure the performance of AI models and systems across various tasks and hardware. 

## 3. Key Performance Metrics:

- **FLOPS (Floating-point Operations Per Second):** Measures the computational power of the GPU. 
- **Memory Bandwidth:** Indicates the speed at which data can be transferred to and from memory. 
- **Latency:** Measures the time it takes for a model to produce a result after receiving an input. 
- **Throughput:** Indicates the number of inferences or training iterations that can be completed per unit of time. 

#### HIP (ROCm) semantics — PyTorch 2.7 documentation

ROCm™ is AMD's open source software platform for GPU-accelerated high performance computing and machine learning. HIP is ROCm's C+

 PyTorch documentation 

#### AMD ROCm™ Software for AI

Optimized for AI Workloads AMD ROCm™ is an open software stack offering a suite of optimizations for AI workloads and supporting t...

 AMD 

#### System health benchmarks - ROCm Documentation - AMD

May 13, 2025 — Before running AI workloads, it is important to validate that your AMD hardware is configured correctly and is perform...

 ROCm Documentation 

#### AMD's AI Chip Strategy: Gaining Ground in a Competitive Market

Sep 16, 2024 — AMD has developed a comprehensive suite of software



In essence, ROCm provides a comprehensive software platform for AI development on AMD hardware, offering tools for benchmarking, performance optimization, and framework integration, ultimately enabling developers to achieve optimal performance and efficiency for their AI workloads. [↗](#)

#### System health benchmarks - ROCm Documentation - AMD

May 13, 2025 — Before running AI workloads, it is important to validate that your AMD hardware is configured correctly and is perform...

[ROCm Documentation](#) [⋮](#)

#### AMD ROCm™ Software for AI

Optimized for AI Workloads AMD ROCm™ is an open software stack offering a suite of optimizations for AI workloads and supporting t...

[AMD](#) [⋮](#)



#### HIP (ROCm) semantics — PyTorch 2.7 documentation

ROCm™ is AMD's open source software platform for GPU-accelerated high performance computing and machine learning. HIP is ROCm's C+

[PyTorch documentation](#) [⋮](#)

Show all



#### HIP (ROCm) semantics — PyTorch 2.7 documentation

ROCm™ is AMD's open source software platform for GPU-accelerated high performance computing and machine learning. HIP is ROCm's C+

[PyTorch documentation](#) [⋮](#)

#### AMD ROCm™ Software for AI

Optimized for AI Workloads AMD ROCm™ is an open software stack offering a suite of optimizations for AI workloads and supporting t...

[AMD](#) [⋮](#)



#### System health benchmarks - ROCm Documentation - AMD

May 13, 2025 — Before running AI workloads, it is important to validate that your AMD hardware is configured correctly and is perform...

[ROCm Documentation](#) [⋮](#)

#### AMD's AI Chip Strategy: Gaining Ground in a Competitive Market

Sep 16, 2024 — AMD has developed a comprehensive suite of software



15 votes · 2 weeks ago

30 votes · 2 weeks ago



See more >



<https://rocm.docs.amd.com> > [how-to](#) > [benchmark-docker](#)

## :

The following models are supported for inference performance **benchmarking** with vLLM and ROCm. Some instructions, commands, and recommendations in this ...

## People also ask



ROCm™ is AMD's open source software platform for GPU-accelerated high performance computing and machine learning. HIP is ROCm's C+



11

Optimized for AI Workloads AMD ROCm™ is an open software stack offering a suite of optimizations for AI workloads and supporting t...



May 13, 2025 — Before running AI workloads, it is important to validate that your AMD hardware is configured correctly and is perform...



Sep 16, 2024 — AMD has developed a comprehensive suite of software



SemiAnalysis

<https://semianalysis.com> > 2025/05/23 > amd-vs-nvidia-...

## AMD vs NVIDIA Inference Benchmark: Who Wins?

May 23, 2025 — Models score worse on various evals on **ROCm** compared to CUDA due to lack of CI and numeric accuracy kernels. SemiAnalysis is hiring additional ...



Tom's Hardware

<https://www.tomshardware.com> > ... > GPUs

## AMD unveils ROCm 7 — new platform boosts AI ...

Jun 13, 2025 — **AMD's Instinct MI300X with ROCm 7** delivers over 3.5 times the inference performance and 3 times the training throughput compared to ROCm 6, according to AMD.



GitHub

<https://github.com> > ROCm > hipBench

## ROCm/hipBench: HIP Kernel Benchmarking Library

Each sample runs the **benchmark** once with a clean device L2 cache. GPU and CPU times are reported.

Batch Measurements: Executes the **benchmark** multiple times back ...

## Videos :

[Maximize AI performance with AMD ROCm tools and Azure ...](#)

49:33

YouTube · Microsoft Developer

May 22, 2025

### HIP (ROCm) semantics — PyTorch 2.7 documentation

ROCm™ is AMD's open source software platform for GPU-accelerated high performance computing and machine learning. HIP is ROCm's C+

[PyTorch documentation](#)

### AMD ROCm™ Software for AI

Optimized for AI Workloads AMD ROCm™ is an open software stack offering a suite of optimizations for AI workloads and supporting t...

[AMD](#)

### System health benchmarks - ROCm Documentation - AMD

May 13, 2025 — Before running AI workloads, it is important to validate that your AMD hardware is configured correctly and is perform...

[ROCm Documentation](#)

### AMD's AI Chip Strategy: Gaining Ground in a Competitive Market

Sep 16, 2024 — AMD has developed a comprehensive suite of software





Reddit · r/LocalLLaMA

50+ comments · 1 month ago

## AMD Takes a Major Leap in Edge AI With ROCm

**AMD Takes a Major Leap in Edge AI With ROCm**; Announces Integration With Strix Halo APUs & Radeon RX 9000 Series GPUs. Resources.

AMD introduces **ROCm 7**, with higher performance ... 52 posts Jun 13, 2025

Enhancing **AI** Training with AMD **ROCm** Software : r ... 12 posts Feb 1, 2025

More results from [www.reddit.com](https://www.reddit.com)

1 2 3 4 Next

70006, Metairie, LA - Based on your past activity - [Update location](#)

[Help](#) [Send feedback](#) [Privacy](#) [Terms](#)



### HIP (ROCm) semantics — PyTorch 2.7 documentation

ROCm™ is AMD's open source software platform for GPU-accelerated high performance computing and machine learning. HIP is ROCm's C+

 [PyTorch documentation](#) ⋮

### AMD ROCm™ Software for AI

Optimized for AI Workloads AMD ROCm™ is an open software stack offering a suite of optimizations for AI workloads and supporting t...

 [AMD](#) ⋮

### System health benchmarks - ROCm Documentation - AMD

May 13, 2025 — Before running AI workloads, it is important to validate that your AMD hardware is configured correctly and is perform...

 [ROCm Documentation](#) ⋮

### AMD's AI Chip Strategy: Gaining Ground in a Competitive Market

Sep 16, 2024 — AMD has developed a comprehensive suite of software

